Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 199 (2017) 1056–1061

Procedia Engineering

www.elsevier.com/locate/procedia

X International Conference on Structural Dynamics, EURODYN 2017

Approximate Bayesian Computation by Subset Simulation for model selection in dynamical systems

Majid K. Vakilzadeh[a,*], James L. Beck[b], Thomas Abrahamsson[a]

[a] *Department of Applied Mechanics, Chalmers University of Technology, Gothenburg, Sweden*
[b] *Division of Engineering and Applied Science, California Institute of Technology, CA, USA*

Abstract

Approximate Bayesian Computation (ABC) methods are originally conceived to expand the horizon of Bayesian inference methods to the range of models for which only forward simulation is available. However, there are well-known limitations of the ABC approach to the Bayesian model selection problem, mainly due to lack of a sufficient summary statistics that work across models. In this paper, we show that formulating the standard ABC posterior distribution as the exact posterior PDF for a hierarchical state-space model class allows us to independently estimate the evidence for each alternative candidate model. We also show that the model evidence is a simple by-product of the ABC-SubSim algorithm. The validity of the proposed approach to ABC model selection is illustrated using simulated data from a three-story shear building with Masing hysteresis.

Peer-review under responsibility of the organizing committee of EURODYN 2017.

Keywords: Approximate Bayesian Computation; Subset Simulation; Bayesian model selection; Masing hysteretic models

1. Introduction

Due to exclusive foundation of Bayesian statistics on the probability logic axioms, it provides a rigorous framework for model updating and model selection. In this approach, a key idea is to construct a *stochastic model class* $\mathcal{M}$ consisting of the following fundamental probability distributions [1]: a set of parameterized input-output probability models $p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{u}, \mathcal{M})$ for predicting the system behavior of interest $\boldsymbol{y}$ for given input $\boldsymbol{u}$ and a *prior* probability density function (PDF) $p(\boldsymbol{\theta}|\mathcal{M})$ over the parameter space $\boldsymbol{\Theta} \in \mathbb{R}^{N_p}$ of $\mathcal{M}$ that reflects the relative degree of plausibility of each input-output model in the set. When data $\mathcal{D}$ consisting of the measured system input $\hat{\boldsymbol{u}}$ and output $\hat{\boldsymbol{z}}$ are available, the prior PDF $p(\boldsymbol{\theta}|\mathcal{M})$ can be updated through Bayes' Theorem to obtain the *posterior* PDF for the uncertain model parameters $\boldsymbol{\theta}$ as:

$$p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}) \propto p(\hat{\boldsymbol{z}}|\boldsymbol{\theta}, \hat{\boldsymbol{u}}, \mathcal{M})p(\boldsymbol{\theta}|\mathcal{M}) \tag{1}$$

where $p(\hat{\boldsymbol{z}}|\boldsymbol{\theta}, \hat{\boldsymbol{u}}, \mathcal{M})$ denotes the *likelihood function* of $\boldsymbol{\theta}$ which gives the probability of getting the data based on the input-output probability model $p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{u}, \mathcal{M})$.

* Corresponding author. Tel.: +46-031-772-1510;
E-mail address: khorsand@chalmers.se

1877-7058
Peer-review under responsibility of the organizing committee of EURODYN 2017.
10.1016/j.proeng.2017.09.291

There are some model classes, e.g., hidden Markov models, for which the likelihood function is difficult or even impossible to compute, but one might still be interested to perform Bayesian parameter inference or model selection. ABC methods were originally conceived to circumvent the need for computation of the likelihood by simulating samples from the corresponding input-output probability model $p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{u}, \mathcal{M})$.

The basic idea behind ABC is to avoid evaluation of the likelihood function in the posterior PDF $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}) \propto p(\hat{\boldsymbol{z}}|\boldsymbol{\theta}, \hat{\boldsymbol{u}}, \mathcal{M})\, p(\boldsymbol{\theta}|\mathcal{M})$ over the parameter space $\boldsymbol{\theta}$ by using an augmented posterior PDF:

$$p(\boldsymbol{\theta}, \boldsymbol{y}|\mathcal{D}, \mathcal{M}) \propto P(\hat{\boldsymbol{z}}|\boldsymbol{y}, \boldsymbol{\theta})\, p(\boldsymbol{y}|\boldsymbol{\theta}, \hat{\boldsymbol{u}}, \mathcal{M})\, p(\boldsymbol{\theta}|\mathcal{M}) \tag{2}$$

over the joint space of the model parameters $\boldsymbol{\theta}$ and the model output $\boldsymbol{y}$ that is simulated using the distribution $p(\boldsymbol{y}|\boldsymbol{\theta}, \hat{\boldsymbol{u}}, \mathcal{M})$. The interesting point of this formulation is the degree of freedom brought by the choice of function $P(\hat{\boldsymbol{z}}|\boldsymbol{y}, \boldsymbol{\theta})$. The original ABC algorithm defines $P(\hat{\boldsymbol{z}}|\boldsymbol{y}, \boldsymbol{\theta}) = \delta_{\hat{\boldsymbol{z}}}(\boldsymbol{y})$, where $\delta_{\hat{\boldsymbol{z}}}(\boldsymbol{y})$ is equal to 1 when $\hat{\boldsymbol{z}} = \boldsymbol{y}$ and equal to 0 otherwise, to retrieve the target posterior distribution when $\boldsymbol{y}$ exactly matches $\hat{\boldsymbol{z}}$. However, the probability of generating exactly $\hat{\boldsymbol{z}} = \boldsymbol{y}$ is zero for continuous stochastic variables. Pitchard et al. [2] broadened the realm of the applications for which ABC algorithm can be used by replacing the point mass at the observed output data $\hat{\boldsymbol{z}}$ with an indicator function $\mathbb{I}_{S(\epsilon)}(\boldsymbol{y})$, where $\mathbb{I}_{S(\epsilon)}(\boldsymbol{y})$ gives 1 over the set $S(\epsilon) = \{\boldsymbol{y} : \rho(\boldsymbol{\eta}(\hat{\boldsymbol{z}}) - \boldsymbol{\eta}(\boldsymbol{y})) \leq \epsilon\}$ and 0 elsewhere, for some chosen metric ρ and low-dimensional summary statistic $\boldsymbol{\eta}$. The ABC algorithm based on this formulation thus gives samples from the true posterior distribution when the tolerance parameter ϵ is sufficiently small and the summary statistics $\boldsymbol{\eta}(.)$ are sufficient. These conditions pose some difficulties for computer implementation of this algorithm which renders it far from a routine use for parameter inference and model selection. Firstly, a sufficiently small tolerance parameter ϵ means that only predicted model outputs $\boldsymbol{y}$ lying in a small local neighborhood centered on the observed data vector $\hat{\boldsymbol{z}}$ are accepted. This leads to the problem of rare-event simulation. To circumvent this problem, Chiachio et al. [3] developed a new algorithm, called ABC-SubSim, by incorporating the Subset Simulation algorithm [4] for rare-event simulation into the ABC algorithm. Secondly, the lack of a reasonable vector of summary statistics that works across models hinders the use of an ABC algorithm for model selection [5].

In this study, we show that formulating a dynamical system as a general hierarchical state-space model enables us to solve the inherent difficulty of the ABC technique to model selection. Using this formulation, one can independently estimate the model evidence for each model class. We also show that the model evidence can be estimated as a simple by-product of the recently proposed multi-level MCMC algorithm, called ABC-SubSim. The effectiveness of the ABC-SubSim algorithm for Bayesian model class selection with simulated data is illustrated using a three-story shear building with Masing hysteresis [6].

2. Formulation

In this section, we review the formulation of a Bayesian hierarchical model class for dynamical systems and then we address the Bayesian model updating and model selection approach for this class of models.

2.1. Formulation of hierarchical stochastic model class

In this section, we present the formulation for a hierarchical stochastic state-space model class $\mathcal{M}$ to predict the uncertain input-output behavior of a system. We start with the general case of a discrete-time finite-dimensional stochastic state-space model of a real dynamic system:

$$\begin{aligned} \forall n \in \mathbb{Z}^+,\ \boldsymbol{x}_n &= \boldsymbol{f}_n(\boldsymbol{x}_{n-1}, \boldsymbol{u}_{n-1}, \boldsymbol{\theta}_s) + \boldsymbol{w}_n \quad \text{(State evolution)} \\ \boldsymbol{y}_n &= \boldsymbol{g}_n(\boldsymbol{x}_n, \boldsymbol{u}_n, \boldsymbol{\theta}_s) + \boldsymbol{v}_n \qquad\qquad \text{(Output)} \end{aligned} \tag{3}$$

where $\boldsymbol{u}_n \in \mathbb{R}^{N_I}$, $\boldsymbol{x}_n \in \mathbb{R}^{N_s}$ and $\boldsymbol{y}_n \in \mathbb{R}^{N_o}$ denote the (external) input, dynamic state and output vector at time t_n, and $\boldsymbol{\theta}_s \in \mathbb{R}^{N_p}$ is a vector of uncertain-valued model parameters. In (3), the uncertain state and output prediction errors $\boldsymbol{w}_n$ and $\boldsymbol{v}_n$ are introduced to account for the model being always an approximation of the real system behavior. The prior distributions, $\mathcal{N}(\boldsymbol{w}_n|\boldsymbol{0}, \boldsymbol{Q}_n(\boldsymbol{\theta}_w))$ and $\mathcal{N}(\boldsymbol{v}_n|\boldsymbol{0}, \boldsymbol{R}_n(\boldsymbol{\theta}_v))$, $\forall n \in \mathbb{Z}^+$, are chosen for the $\boldsymbol{w}_n$ and $\boldsymbol{v}_n$ based on the Principle of Maximum (Information) Entropy [7], where $\{\boldsymbol{w}_n\}_{n=1}^N$ and $\{\boldsymbol{v}_n\}_{n=1}^N$ are sequences of independent stochastic variables. We add the uncertain parameters that specify these priors to the model parameters $\boldsymbol{\theta}_s$ and use $\boldsymbol{\theta} = [\boldsymbol{\theta}_s^T\ \boldsymbol{\theta}_w^T\ \boldsymbol{\theta}_v^T]^T$ to denote the uncertain parameter vector for the stochastic state-space model. Then, we choose a prior $p(\boldsymbol{\theta}|\mathcal{M})$ for all of the model class parameters.

The defined stochastic state-space model defines a "hidden" Markov chain for the state time history $\{\boldsymbol{x}\}_{n=1}^N$ (which will also be denoted by the vector $\boldsymbol{x}_{1:N} = [\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_N^T]^T \in \mathbb{R}^{NN_s}$) by implying a state transition PDF:

$$\forall n \in \mathbb{Z}^+,\ p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1}, \boldsymbol{u}_{n-1}, \boldsymbol{\theta}, \mathcal{M}) = \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{f}_n(\boldsymbol{x}_{n-1}, \boldsymbol{u}_{n-1}, \boldsymbol{\theta}), \boldsymbol{Q}_n(\boldsymbol{\theta})) \tag{4}$$

along with a state-to-output PDF:

$$\forall n \in \mathbb{Z}^+,\ p(\boldsymbol{y}_n|\boldsymbol{x}_n, \boldsymbol{u}_n, \boldsymbol{\theta}, \mathcal{M}) = \mathcal{N}(\boldsymbol{y}_n|\boldsymbol{g}_n(\boldsymbol{x}_n, \boldsymbol{u}_n, \boldsymbol{\theta}), \boldsymbol{R}_n(\boldsymbol{\theta})) \tag{5}$$

These, in turn, imply the following two probability models:

$$p(\boldsymbol{x}_{1:N}|\boldsymbol{u}_{0:N}, \boldsymbol{\theta}, \mathcal{M}) = \prod_{n=1}^N p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1}, \boldsymbol{u}_{n-1}, \boldsymbol{\theta}, \mathcal{M}) \tag{6}$$

$$p(\boldsymbol{y}_{1:N}|\boldsymbol{x}_{1:N}, \boldsymbol{u}_{0:N}, \boldsymbol{\theta}, \mathcal{M}) = \prod_{n=1}^N p(\boldsymbol{y}_n|\boldsymbol{x}_n, \boldsymbol{u}_n, \boldsymbol{\theta}, \mathcal{M}) \tag{7}$$

The stochastic input-output model (or forward model) for given parameter vector $\boldsymbol{\theta}$ is then:

$$p(\boldsymbol{y}_{1:N}|\boldsymbol{u}_{0:N}, \boldsymbol{\theta}, \mathcal{M}) = \int p(\boldsymbol{y}_{1:N}|\boldsymbol{x}_{1:N}, \boldsymbol{u}_{0:N}, \boldsymbol{\theta}, \mathcal{M}) p(\boldsymbol{x}_{1:N}|\boldsymbol{u}_{0:N}, \boldsymbol{\theta}, \mathcal{M}) d\boldsymbol{x}_{1:N} \tag{8}$$

This high-dimensional integral usually cannot be done analytically. We will therefore structure the stochastic input-output model using a Bayesian hierarchical model to avoid the integration in (8).

This can be done by extending the stochastic model class $\mathcal{M}$ to a new one $\mathcal{M}(\epsilon)$ that also predicts the measured system output $\boldsymbol{z}_n$ at time t_n:

$$\boldsymbol{z}_n = \boldsymbol{y}_n + \boldsymbol{e}_n = \boldsymbol{g}_n(\boldsymbol{x}_n, \boldsymbol{u}_n, \boldsymbol{\theta}) + \boldsymbol{v}_n + \boldsymbol{e}_n \tag{9}$$

where $\boldsymbol{e}_n$ denotes the uncertain measurement error at time t_n. Here, we allow for dependence between the set of stochastic variables $\{\boldsymbol{e}_n\}_{n=1}^N$ by specifying a joint PDF for $\boldsymbol{e}_{1:n}$ for any $n \in \mathbb{Z}^+$. A simple choice for the probability model for the measurement error $\boldsymbol{e}_{1:n}$ is a uniform PDF which gives the following predictive PDF for the observed system output (sensor output) $\boldsymbol{z}_{1:n}$ conditioned on the actual system output $\boldsymbol{y}_{1:n}$:

$$p(\boldsymbol{z}_{1:n}|\boldsymbol{y}_{1:n}, \mathcal{M}(\epsilon)) = p(\boldsymbol{e}_{1:n}|\mathcal{M}(\epsilon))\Big|_{\boldsymbol{e}_{1:n}=\boldsymbol{z}_{1:n}-\boldsymbol{y}_{1:n}} = \begin{cases} V_n(\epsilon)^{-1} & \text{if } \|\boldsymbol{z}_{1:n} - \boldsymbol{y}_{1:n}\| \leq \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where $V_n(\epsilon) = \int_{\mathbb{R}^{N_o n}} \mathbb{I}_\epsilon(\boldsymbol{e}_{1:n})\, d\boldsymbol{e}_{1:n}$ is the volume of region $S(\epsilon)$, and $\mathbb{I}_\epsilon(\boldsymbol{e}_{1:n})$ is the indicator function for the set $S(\epsilon) = \{\boldsymbol{e}_{1:n} \in \mathbb{R}^{N_o n} : \|\boldsymbol{e}_{1:n}\| \leq \epsilon\}$ for some vector norm on $\mathbb{R}^{N_o n}$. Now, the specification of the hierarchical prior PDF based on (6) and (7):

$$p(\boldsymbol{y}_{1:N}, \boldsymbol{x}_{1:N}, \boldsymbol{\theta}|\boldsymbol{u}_{0:N}, \mathcal{M}(\epsilon)) = p(\boldsymbol{y}_{1:N}|\boldsymbol{x}_{1:N}, \boldsymbol{u}_{0:N}, \boldsymbol{\theta}, \mathcal{M}(\epsilon)) p(\boldsymbol{x}_{1:N}|\boldsymbol{u}_{0:N}, \boldsymbol{\theta}, \mathcal{M}(\epsilon))\, p(\boldsymbol{\theta}|\mathcal{M}(\epsilon)) \tag{11}$$

completes the definition of the stochastic model class $\mathcal{M}(\epsilon)$ that is the PDFs in (10) and (11).

2.2. Bayesian model updating

If measured system input and system output data, $\mathcal{D}_N = \{\hat{\boldsymbol{u}}_{0:N}, \hat{\boldsymbol{z}}_{1:N}\}$, are available from the dynamic system, then the predictive PDF in (10) with $n = N$ gives the likelihood function:

$$p(\hat{\boldsymbol{z}}_{1:N}|\boldsymbol{y}_{1:N}, \mathcal{M}(\epsilon)) = \frac{\mathbb{I}_{\mathcal{D}(\epsilon)}(\boldsymbol{y}_{1:N})}{V_N(\epsilon)} \tag{12}$$

with the indicator function defined over the set $\mathcal{D}(\epsilon) = \{\boldsymbol{y}_{1:N} \in \mathbb{R}^{NN_o} : \|\boldsymbol{y}_{1:N} - \hat{\boldsymbol{z}}_{1:N}\| < \epsilon\}$, where $\|.\|$ is some vector norm on $\mathbb{R}^{NN_o}$. The posterior PDF for stochastic model class $\mathcal{M}(\epsilon)$ is then given by Bayes' Theorem:

$$p(\boldsymbol{y}_{1:N}, \boldsymbol{x}_{1:N}, \boldsymbol{\theta}|\mathcal{D}_N, \mathcal{M}(\epsilon)) = E(\epsilon)^{-1} \frac{\mathbb{I}_{\mathcal{D}(\epsilon)}(\boldsymbol{y}_{1:N})}{V_N(\epsilon)} p(\boldsymbol{y}_{1:N}, \boldsymbol{x}_{1:N}, \boldsymbol{\theta}|\hat{\boldsymbol{u}}_{0:N}, \mathcal{M}(\epsilon)) \tag{13}$$

where the evidence for $\mathcal{M}(\epsilon)$ is then defined as:

$$E(\epsilon) = p(\hat{\boldsymbol{z}}_{1:N}|\hat{\boldsymbol{u}}_{0:N}, \mathcal{M}(\epsilon)) = \int p(\hat{\boldsymbol{z}}_{1:N}|\boldsymbol{y}_{1:N}, \epsilon)\, p(\boldsymbol{y}_{1:N}, \boldsymbol{x}_{1:N}, \boldsymbol{\theta}|\hat{\boldsymbol{u}}_{0:N}, \mathcal{M}(\epsilon))\, d\boldsymbol{y}_{1:N}\, d\boldsymbol{x}_{1:N}\, d\boldsymbol{\theta} \tag{14}$$

The theory for the hierarchical model and its updating presented so far in Section 2 is general and valid for any $\epsilon > 0$ deemed appropriate. For the application of ABC, we suppose that $\mathcal{M}(0) \equiv \mathcal{M}(\epsilon \to 0)$ is actually the stochastic model class of interest. For ϵ sufficiently small, the set $\mathcal{D}(\epsilon)$ of outputs $\boldsymbol{y}_{1:N}$ will converge to the observed output vector $\hat{\boldsymbol{z}}_{1:N}$ and the posterior PDF in (13) for stochastic model class $\mathcal{M}(\epsilon)$ will converge to the desired posterior distribution of the model parameters $p(\boldsymbol{\theta}|\mathcal{D}_N, \mathcal{M}(0))$ after marginalization.

Remark. Vakilzadeh et al. [8] showed that for the hierarchical stochastic model class $\mathcal{M}(\epsilon)$, the exact posterior PDF (13) using a uniformly-distributed uncertain measurement error in the output space is identical to the ABC posterior PDF given for no measurement error. Thus, ABC-SubSim that was originally developed by Chiachio et al. [3] to draw samples from an ABC posterior PDF can be used to solve the exact Bayesian problem for the hierarchical stochastic model class $\mathcal{M}(\epsilon)$.

2.3. Bayesian model class assessment

Consider a set $\boldsymbol{M} \equiv \{\mathcal{M}_1(\epsilon_{\mathcal{M}_1}), \mathcal{M}_2(\epsilon_{\mathcal{M}_2}), \ldots, \mathcal{M}_L(\epsilon_{\mathcal{M}_L})\}$ of L Bayesian hierarchical model classes for representing a system. In Bayesian model selection, models in $\boldsymbol{M}$ are ranked based on their probabilities conditioned on the data $\mathcal{D}_N$ that is given by Bayes' Theorem:

$$P(\mathcal{M}_j(\epsilon_{\mathcal{M}_j})|\mathcal{D}_N) = \frac{p(\hat{\boldsymbol{z}}_{1:N}|\hat{\boldsymbol{u}}_{0:N}, \mathcal{M}_j(\epsilon_{\mathcal{M}_j}))P(\mathcal{M}_j(\epsilon_{\mathcal{M}_j})|\boldsymbol{M})}{\sum_{l=1}^{L} p(\hat{\boldsymbol{z}}_{1:N}|\hat{\boldsymbol{u}}_{0:N}, \mathcal{M}_j(\epsilon_{\mathcal{M}_j}))P(\mathcal{M}_j(\epsilon_{\mathcal{M}_j})|\boldsymbol{M})} \tag{15}$$

where $P(\mathcal{M}_j(\epsilon_{\mathcal{M}_j})|\boldsymbol{M})$ denotes the prior probability of $\mathcal{M}_j(\epsilon_{\mathcal{M}_j})$ that indicates the modeler's belief about the initial relative plausibility of $\mathcal{M}_j(\epsilon_{\mathcal{M}_j})$ within the set $\boldsymbol{M}$. The factor $p(\hat{\boldsymbol{z}}_{1:N}|\hat{\boldsymbol{u}}_{0:N}, \mathcal{M}_j(\epsilon_{\mathcal{M}_j}))$, which is the evidence (or marginal likelihood) for $\mathcal{M}_j(\epsilon_{\mathcal{M}_j})$, indicates the probability of data $\mathcal{D}_N$ according to $\mathcal{M}_j(\epsilon_{\mathcal{M}_j})$.

For the specific choice of Bayesian hierarchical model class, the evidence can be estimated by (14). However, its calculation requires the evaluation of a high-dimensional integral which is the computationally challenging step in Bayesian model selection, especially as $\epsilon \to 0$. ABC-SubSim provides a straightforward approximation for it via the conditional probabilities involved in the Subset Simulation. Indeed, the last integral in (14) is the probability $P(\boldsymbol{y}_{1:N} \in \mathcal{D}(\epsilon_{\mathcal{M}_j})|\mathcal{M}_j)$ that $\boldsymbol{y}_{1:N}$ belongs to $\mathcal{D}(\epsilon_{\mathcal{M}_j}) = \{\boldsymbol{y}_{1:N} \in \mathbb{R}^{NN_o} : \|\boldsymbol{y}_{1:N} - \hat{\boldsymbol{z}}_{1:N}\| \leq \epsilon_{\mathcal{M}_j}\}$. This probability can be readily estimated as a by-product of ABC-SubSim. Thus, for a particular tolerance level $\epsilon_{\mathcal{M}_j}$ and model class $\mathcal{M}_j(\epsilon_{\mathcal{M}_j})$, the evidence is estimated by:

$$\hat{E}_{\mathcal{M}_j} = \frac{P(\boldsymbol{y}_{1:N} \in \mathcal{D}(\epsilon_{\mathcal{M}_j})|\mathcal{M}_j)}{V_N(\epsilon_{\mathcal{M}_j})} = \frac{1}{V_N(\epsilon_{\mathcal{M}_j})} P_0^{i-1}\, P_i \tag{16}$$

where i would be such that $\epsilon_i \leq \epsilon_{\mathcal{M}_j} < \epsilon_{i-1}$, in which the intermediate "radii" ϵ_i's are automatically chosen by ABC-SubSim, P_i is the fraction of samples generated in $\mathcal{D}(\epsilon_{i-1})$ that lie in $\mathcal{D}(\epsilon_{\mathcal{M}_j})$, and P_0 is the selected conditional probability at each simulation level of ABC-SubSim.

Wilkinson [9] also showed that a standard ABC posterior gives an exact posterior distribution for a new model under the assumption that the summary statistics are corrupted with a uniform additive error term. However, formulating standard ABC based on summary statistics hinders the independent approximation of evidence for each candidate model. Here, we provided an estimate of the model evidence in (16) as a result of formulating a dynamic problem in terms of a general hierarchical stochastic state-space model where the likelihood function $p(\hat{\boldsymbol{z}}_{1:N}|\boldsymbol{y}_{1:N}, \mathcal{M}(\epsilon))$ is expressed using the entire data $\hat{\boldsymbol{z}}_{1:N}$ and ABC-SubSim readily produces an unbiased approximation of the evidence.

3. Illustrative Example: Three-story Masing shear-building under seismic excitation

This example, which is taken from Muto and Beck [6], considers a three-story shear building with the following equation of motion:

$$\boldsymbol{M}\ddot{\boldsymbol{z}}(t) + \boldsymbol{C}\dot{\boldsymbol{z}}(t) + \boldsymbol{f}_h = -\boldsymbol{M}\mathbf{1}u(t) \tag{17}$$

where $\boldsymbol{z}(t) \in \mathbb{R}^3$ is the horizontal displacement vector relative to the ground; $\boldsymbol{M}$, $\boldsymbol{C} \in \mathbb{R}^{3\times 3}$ are the mass and damping matrices; $u(t)$ is the horizontal ground acceleration; and $\mathbf{1} = [1\ 1\ 1]^T$. The restoring force for the ith story is given by $f_{h,i} = r_i - r_{i+1}$ where the inter-story shear force-deflection relation r_i is given by the *Masing hysteretic model* [10]. An interested reader is referred to [10] for a detailed description of the Masing hysteretic model. r_i for each story can be characterized by three parameters: small-amplitude inter-story stiffnesses k_i, ultimate strength $r_{u,i}$, and elastic-to-plastic transition parameter α_i.

In this example, the structure has a known story mass of 1.25×10^5 kg. The viscous damping matrix $\boldsymbol{C}$ is modeled using Rayleigh damping $\boldsymbol{C} = c_M\boldsymbol{M} + c_K\boldsymbol{K}$. The actual values for the model parameters for the hysteresis model and the viscous damping matrix are presented in Table 1. The prior distribution over the nine-dimensional parameter space of the hysteresis model is selected to be the product of nine lognormal PDFs with logarithmic mean value of $\log(2.5 \times 10^8)$, $\log(2.5 \times 10^6)$, and $\log(4)$ for k_i , $r_{u,i}$, and α_i, $i = 1, 2, 3$, respectively, and a logarithmic standard deviation of 0.5 for all of them. The prior PDFs for the parameters of the damping matrix are defined as independent uniform PDFs over the interval $[0, 1.5]$ for c_M and $[0, 1.5 \times 10^{-3}]$ for c_K. The east-west component of the Sylmar ground-motion record from the County Hospital Parking Lot during 1994 Northridge earthquake in California is used here as the excitation. The synthetic response data for system identification is the inter-story drift time histories from the oracle model, for which the parameters are set to their actual values, and the standard deviation of the uncertain output error is set to 0.03 cm. For ABC-SubSim, the number of samples in each level is fixed to $N_t = 2000$, the adaptation probability to $P_a = 0.1$, and the conditional probability to $P_0 = 0.1$.

Four model classes are studied for system identification. For model classes $\mathcal{M}_1$ and $\mathcal{M}_2$, the elastic-to-plastic transition parameters are constrained to be equal for all three stories whereas they are allowed to vary for model classes $\mathcal{M}_3$ and $\mathcal{M}_4$. The model classes $\mathcal{M}_1$ and $\mathcal{M}_3$ contain no viscous damping, but model classes $\mathcal{M}_2$ and $\mathcal{M}_4$ do.

Table 1 shows the MAP (maximum a posteriori) values and the standard deviations of the uncertain parameters obtained for all model classes. Table 2 shows the final tolerance levels $\epsilon_{\mathcal{M}_j}$ for different model classes. This table also presents the posterior probability of model classes $P(\mathcal{M}_j(\epsilon_{\mathcal{M}_j})|\mathcal{D}_N, \boldsymbol{M})$, $j = 1, 2, 3, 4$ calculated by evaluation of evidence (15) at the final tolerance levels $\epsilon_{\mathcal{M}_j}$ and equal prior probabilities $P(\mathcal{M}_j|\boldsymbol{M}) = 1/4$ for the models. It is not surprising that the posterior probability for the model classes favors model class $\mathcal{M}_2$ since it contains the model used to generate the synthetic data and has two parameters less than model class $\mathcal{M}_4$, which also contains the data-generating model. As shown by the information-theoretic expression for the log evidence in [6], the posterior probability of a model class is controlled by a trade-off between the posterior average data fit (the posterior mean of the log-likelihood) and the amount of information extracted from data (the relative entropy of the posterior with respect to the prior). $\mathcal{M}_2$ and $\mathcal{M}_4$ give essentially the same average data fit but $\mathcal{M}_2$ extracts less information abouts its parameters.

The approximate posterior probabilities $P(\mathcal{M}_j(\epsilon_{\mathcal{M}_j})|\mathcal{D}_N, \boldsymbol{M})$ presented in Table 2 are in agreement with those reported by Muto and Beck [6]. Figures 1 (Left and Right), respectively show the probability that $\boldsymbol{y}_{1:N}$ falls in the data-approximating region $\mathcal{D}(\epsilon)$ and the posterior probability $P(\mathcal{M}_j(\epsilon)|\mathcal{D}_N, \boldsymbol{M})$ for different model classes versus the tolerance level ϵ.

4. Concluding remarks

In the current state of the art, ABC methods can only be used for model class selection in a very limited range of models for which a set of sufficient summary statistics can be found so that it also guarantees sufficiency across the set of models under study. In this paper, a new ABC model selection procedure has been presented which broadens the realm of ABC-based model comparison to be able to assess dynamic models. The presented numerical example showed the effectiveness of the proposed method for ABC model selection.

Table 1 The MAP values and the standard deviations (in parentheses) of the uncertain model parameters.

Model class	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	Oracle model
k_1 (10^8 N/m)	2.586 (0.026)	2.497 (0.018)	2.545 (0.029)	2.490 (0.015)	2.500
k_2 (10^8 N/m)	2.455 (0.044)	2.499 (0.025)	2.539 (0.042)	2.509 (0.034)	2.500
k_3 (10^8 N/m)	2.566 (0.054)	2.490 (0.023)	2.545 (0.061)	2.504 (0.025)	2.500
$r_{u,1}$ (10^6 N)	1.737 (0.004)	1.749 (0.003)	1.746 (0.006)	1.751 (0.003)	1.750
$r_{u,2}$ (10^6 N)	1.779 (0.064)	1.750 (0.037)	1.924 (0.152)	1.757 (0.054)	1.750
$r_{u,3}$ (10^6 N)	2.056 (1.014)	2.140 (0.772)	2.358 (0.934)	2.154 (1.083)	1.750
α_1	3.430 (0.090)	3.981 (0.075)	3.447 (0.145)	4.041 (0.094)	4
α_2	$= \alpha_1$	$= \alpha_1$	2.626 (0.300)	3.863 (0.411)	4
α_3	$= \alpha_1$	$= \alpha_1$	2.552 (2.607)	3.332 (2.047)	4
$c_M(s^{-1})$	—	0.259 (0.071)	—	0.283 (0.069)	0.293
$c_K(10^{-4}s)$	—	2.295 (1.322)	—	2.116 (0.909)	2.640
$\sigma_v(10^{-4}m)$	5.293 (0.063)	3.197 (0.064)	5.163 (0.084)	3.176 (0.048)	3.000

Table 2 Posterior probability of different model classes together with final tolerance level.

Model class	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
Tolerance level ($\epsilon_{\mathcal{M}_j}$)	6.80×10^{-4}	4.25×10^{-4}	7.10×10^{-4}	4.25×10^{-4}
$P(\mathcal{M}_j(\epsilon_{\mathcal{M}_j})\|\mathcal{D}_N, \boldsymbol{M})$	0	0.982	0	0.018

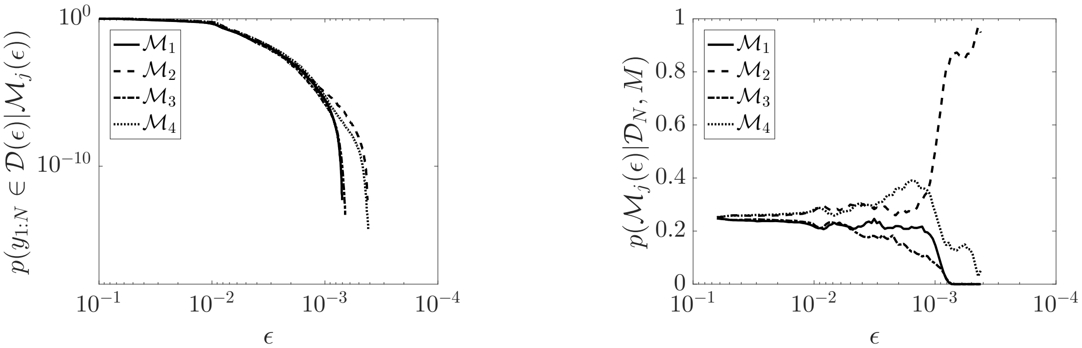

Fig. 1 *Left)The probability of entering the data-approximating region $\mathcal{D}(\epsilon)$ against tolerance level ϵ; Right) The posterior probability of different model classes $\mathcal{M}_j$ against tolerance level ϵ.*

References

[1] J. L. Beck, Bayesian system identification based on probability logic, Structural Control and Health Monitoring 17 (2010) 825–847.

[2] J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaun, M. W. Feldman, Population growth of human Y chromosomes: a study of Y chromosome microsatellites., Molecular Biology and Evolution 16 (1999) 1791–1798.

[3] M. Chiachio, J. L. Beck, J. Chiachio, G. Rus, Approximate Bayesian computation by Subset Simulation, SIAM Journal on Scientific Computing 36 (2014) A1339–A1358.

[4] S.-K. Au, J. L. Beck, Estimation of small failure probabilities in high dimensions by Subset Simulation, Probabilistic Engineering Mechanics 16 (2001) 263–277.

[5] X. Didelot, R. G. Everitt, A. M. Johansen, D. J. Lawson, et al., Likelihood-free estimation of model evidence, Bayesian Analysis 6 (2011) 49–76.

[6] M. Muto, J. L. Beck, Bayesian updating and model class selection for hysteretic structural models using stochastic simulation, Journal of Vibration and Control 14 (2008) 7–34.

[7] E. T. Jaynes, Information theory and statistical mechanics, Physical Review 106 (1957) 620.

[8] M. K. Vakilzadeh, Y. Huang, J. L. Beck, T. Abrahamsson, Approximate bayesian computation by subset simulation using hierarchical state-space models, Mechanical Systems and Signal Processing 84, Part B (2017) 2 – 20.

[9] R. D. Wilkinson, Approximate Bayesian Computation (ABC) gives exact results under the assumption of model error, Statistical Applications in Genetics and Molecular Biology 12 (2013) 129–141.

[10] P. Jayakumar, J. L. Beck, System identification using nonlinear structural models, in: Structural safety evaluation based on system identification approaches, Springer, 1988, pp. 82–102.